Pixel-Aware Stable Diffusion for Realistic Image Super-resolution and Personalized Stylization

Abstract

Diffusion models have demonstrated impressive performance in various image generation, editing, enhancement and translation tasks. In particular, the pre-trained text-to-image stable diffusion models provide a potential solution to the challenging realistic image super-resolution (Real-ISR) and image stylization problems with their strong generative priors. However, the existing methods along this line often fail to keep faithful pixel-wise image structures. If extra skip connections between the encoder and the decoder of a VAE are used to reproduce details, additional training in image space will be required, limiting the application to tasks in latent space such as image stylization. In this work, we propose a pixel-aware stable diffusion (PASD) network to achieve robust Real-ISR and personalized image stylization. Specifically, a pixel-aware cross attention module is introduced to enable diffusion models perceiving image local structures in pixel-wise level, while a degradation removal module is used to extract degradation insensitive features to guide the diffusion process together with image high level information. An adjustable noise schedule is introduced to further improve the image restoration results. By simply replacing the base diffusion model with a stylized one, PASD can generate diverse stylized images without collecting pairwise training data, and by shifting the base model with an aesthetic one, PASD can bring old photos back to life. Extensive experiments in a variety of image enhancement and stylization tasks demonstrate the effectiveness of our proposed PASD approach. Our source codes are available at \url{https://github.com/yangxy/PASD/}.